NASA TECHNICAL MEMORANDUM

NASA TM X-53503

August 15, 1966

NASA TM X-53503

GPO PRICE $ ______

CFSTI PRICE(S) $ ______

Hard copy (HC) 1.00

Microfiche (MF) .50

ff 653 July 65

INFORMATION SCIENCE

By VLADIMIR SLAMECKA

N66 37803	
(ACCESSION NUMBER)	(THRU)
18	1
(PAGES)	(CODE)
TMX-53503	34
(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)

FACILITY FORM 602

NASA

George C. Marshall Space Flight Center, Huntsville, Alabama

NASA - GEORGE C. MARCHALL SPACE FLIGHT CENTER

TECHNICAL MEMORANDUM X-53503

INFORMATION SCIENCE

By

Dr. Vladimir Slamecka *

* Georgia Institute of Technology

RESEARCH PROJECTS LABORATORY
RESEARCH AND DEVELOPMENT OPERATIONS

TECHNICAL MEMORANDUM X-53503

INFORMATION SCIENCE

By

Dr. Vladimir Slamecka

ABSTRACT

This presentation discusses some general and less obvious characteristics of man's knowledge, as we understand it, and some of the problems which we encounter when we seek to design ways of controlling it. The information is presented from the viewpoint of a newer field of scientific and engineering endeavor - that of information science. The discussion appropriately describes this framework of reference, its content, and its tools. Two topics are discussed: the phenomenology and methods of information science; and, more selectively, the processes of knowledge generation, growth and control.

NASA-GEORGE C. MARSHALL SPACE FLIGHT CENTER

BIOGRAPHICAL NOTE

Dr. Vladimir Slamecka is Director, School of Information Science, Georgia Institute of Technology, Atlanta, Georgia. He earned his B. S. degree in chemical engineering at the Benes University of Technology in Brno, Czechoslovakia. He took postgraduate studies at the Faculty of Science, University of Sydney, Australia, and at the University of Munich, Germany. He received a Master's degree in Library Science from Columbia University in 1958 and a Doctorate in Library Science from Columbia in 1962.

Dr. Slamecka was engaged in private chemical engineering research 1955-56. He taught at the School of Engineering, Columbia University, 1958-60. He received a grant from the National Science Foundation to survey the organization, current plans, and communication networks of scientific research of selected East European countries in 1961.

Following graduation from Columbia in 1962, Dr. Slamecka joined Documentation Incorporated as Manager of the Special Studies Division. In 1963, he was appointed as their Manager of the Cancer Chemotherapy National Service Center (CCNSC) data processing activities. Dr. Slamecka was appointed to his present position in 1964.

Dr. Slamecka has contributed to the design of large information systems; information technology; man-machine techniques; applications of information theory to file organization; and the methodology of the discipline. As manager of the CCNSC information activities, he had administrative and technical responsibility for the largest electronic information processing system in the biomedical field.

Dr. Slamecka's present work and interests include the development and administration of a graduate program in information science. This encompasses a study of the phenomenology and theoretical fundamentals underlying this discipline and its various component areas such as computer science documentation, cybernetics, linguistics analysis, and related topics.

TABLE OF CONTENTS

TECHNICAL MEMORANDUM X-53503

INFORMATION SCIENCE*

SUMMARY

This presentation discusses some general and less obvious characteristics of man's knowledge, as we understand it, and some of the problems which we encounter when we seek to design ways of controlling it. The information is presented from the viewpoint of a newer field of scientific and engineering endeavor - that of information scienoe. The discussion appropriately describes this framework of reference, its content, and its tools. Two topics are discussed: the phenomenology and methods of information science; and, more selectively, the processes of knowledge generation, growth and control.

THE PHENOMENOLOGY OF INFORMATION SCIENCE

I am frequently asked to explain, or define, the field of information science. When I answer with what I think is judicious deliberation but what must appear as fumbling with a fuzzy concept, the inquirer quickly takes the upper hand and declares boldly, "How do you know there is such a science, anyway?" or, "How do you know it is a science?"

Whether or not information science is a science is immaterial, at least at this stage. However, an inquiry of this nature helps to throw some light on the process of the coming about of a scientific discipline or area, and is therefore not entirely futile.

There is some evidence that scientific areas behave somewhat as living organisms; they have their birth, and their moments of truth and decay [1]. There is also some evidence that the birth of such a field observes a certain sequence of events. In the very beginning, there usually exists some need - a social, sociological, economic, or technological pressure which gives rise to the first attempts to deal with or satisfy this need. Out of these attempts develops in time a cluster of skills - a profession. The third phase begins with endeavors to give these skills a "raison d'être," a quantitative scientific basis. This last phase signals the birth of a scientific area or discipline.

---

* This material was presented at a Space Science Seminar at MSFC, May 19, 1966.

A favorite example demonstrating the science birth algorithm is that of medicine. Hundreds and even thousands of years ago, the social need - to alleviate pain and to cure the sick - was in the province of the medicine man. He had his counterpart among the American Indians, ancient Egyptians, Australian aborigines, etc. Only fifty or less years ago, the physician in the Mid-East was frequently the barber, opening veins and pulling out worms. In central Europe, the local blacksmith was also the dentist. The educated physician of today is a professional man, not a scientist; and it is only in the last decade or two that we are beginning to speak of a "medical science," and to develop and apply the scientific method in the investigation of health phenomena. Parenthetically, the introduction of the scientific method into medicine is probably one of the first emerging significant contributions of information science.

The span of time leading to the beginning of a science of health has occupied a dozen or more centuries. Clearly this span of time may vary for other disciplines, and it may depend on many factors. One wonders, however, what is the significance of the argument which one might make - that the span of time leading to the development of information science has been less than one quarter of a century.

In the instance of information science, the three phases of need, skills and theory followed so closely that they nearly coincided. Immediately after World War II, there arose an acute awareness of the information "explosion," a phrase which connotes a deficiency in making decisions optimally and in solving problems rapidly; problems such as that of managing soundly growing corporations and industries, or that of closing the space handicap at the time of the first Sputnik. Nearly concurrent with the suddenly recognized need came an advance in electronic technology, the development and production of high-speed information processing devices which devour and operate on symbols at a rate one billion times faster than man. This led to the devising, by trial and error, of techniques and skills (programming) a means by which the computer can be applied to handle information. The third phase, that of the beginnings of theoretical foundations, may be exemplified by the outgrowths of Shannon's communication theory.

Thus within the past twenty years, there have occurred a pressing need, a technological and professional "know-how," and even a beginning of theoretical foundations to give birth to a new field and aspiring discipline, information science.

For a field to aspire to become a scientific discipline it should have the three characteristics which follow.

1. First, it should be concerned with the study of some phenomenon of man, nature or the universe. In our case, this phenomenon is "information;" information science then is concerned with the study of the nature and properties of information, in much the same way in which physics is concerned with the study of energy. The parallel between information science and physics is more than superficial. For example, there appears to be a formal relationship between information theory and thermodynamics, a relationship which awaits to be explored and exploited by information science. Despite the current differing frameworks of the two areas - the one subjective, the other objective, a hope exists that the body of knowledge in thermodynamics can be brought to bear on information science and engineering.

What is the nature of information? Is it, as it appears to be for some of its forms, a property of a particular configuration or structure of elements? This appears to hold for the genetic form of information contained in the molecule of DNA. Or is it a "commodity," as some would regard it, or a lifeblood in certain processes of communication? Or is it all of these? We do not quite know, and may not know for some time to come. The persistent attempts at an encompassing definition are perhaps futile attempts at a "reductio ad absurdum," scientifically as meaningless as the "definition" by the radio announcer overheard recently: "Information is the hallmark of democracy." The difficulty to define basic phenomena of nature uniquely and encompassingly is certainly not new.

In general, science then resorts to a study of the properties of such a phenomenon, and through it seeks to understand the nature of the phenomenon itself. If information is a "property" of a certain configuration of symbols, then the study of the way in which symbols "represent" information may be expected to shed some light on the nature of information. On a less metaphysical level, the properties of symbols and symbol languages - numeric, alphabetic, chemical, etc. - very much determine the efficiency and potential of their use in communication in the real world.

Because at least some forms of information have the apparently subjective property of meaning, the process of representing concepts by symbols is not a generally reversible one, and hence difficult to study. The difficulties are greater when we attempt to study the structural properties of information or symbols - the manner in which an element of information or a symbol is related to another element or symbol. The study of information relations and the structures reflecting these relations lies at the very basis of both the classification and manipulation of information.

The study of information representation, structures, and transformations is at present the main direction of effort in information science aimed toward understanding the nature and properties of information. Dependent on its results is the exploitation of much of the computer's potential.

2. Secondly, the aspiring discipline should be concerned with some processes in man, nature, or universe in which this phenomenon occurs.

In a narrower sense, information science is concerned with information processing, its techniques and devices. Information processing techniques pertain to processes such as information generation, description, categorization, storage, deployment and transformation - all of which depend on the representation and structuring properties of information and symbols. Information processing devices subsume both natural (e.g., human brain) and artificial devices (e.g., computer), communication (e.g., book) as well as storage devices (e.g., film), etc. Information science studies these devices as information processors.

On a higher level, information science shares with and contributes to cybernetics in the study of organization, that is, of the structure within which information operates and of the changes in the states of this structure which information brings about. The processes of learning, adaptation, and self-organization in both natural and artificial systems are, when approached from the viewpoint of information processing, within the realm of study of information science.

3. Thirdly, the aspiring discipline should have a means of proof and feedback through application. This activity of information science (or better, information engineering), is concerned with the design, implementation and evaluation of information systems and their components. At the present time, information engineering aids mainly in problem solving, decision making, communication and learning. It is these applications which are most familiar to the observer: management information systems, document storage/retrieval systems, medical information systems, economic system models, command and control systems, scientific computing systems, and general user-oriented computing facilities. On the horizon are information communication networks, information-based (computer-aided) educational systems, and a variety of nation-wide information networks with a growing repertoire of problem solving aids built into them. These are the building blocks of the new world in which information science and technology promise to extend significantly the power and potential of man's intellect. I believe, however, that this promise will be fully realized only through a new insight into the nature and fundamental properties of information.

Information science is the study of the nature and properties of information, and of information processing techniques and devices. It finds engineering applications in the design of information systems and their components serving a broad variety of purposes. Incidentally, this concept of information science is parallel to or subsumes fields such as "communication science" (of the University of Michigan) and "computer science," and it is nearly identical with the European and Soviet understanding of "cybernetics."

Briefly, what are the intellectual tools of information science? While developing some of its own, most of the ones available now come from traditional and newer disciplines or fields; their power lies in their concomitant multidisciplinary use. I shall enumerate some of these. In information science, some of the techniques having a more pronounced effect are: from mathematics, mathematical analysis, modern algebra, probability and statistics, decision theory and other specialized areas; from philosophy, logics and epistemology; from linguistics, its intersection with logic and mathematics (known as semiotics, or the study of symbols). Since information is man-related, behavioral sciences (physiology, phychology and bionics) are most relevant. With respect to engineering applications, such applied mathematical tools as operations research and decision theory are used frequently. General systems theory promises to be a cornerstone of information engineering.

ON KNOWLEDGE AND ITS CONTROL

I should now like to discuss a few aspects of the body of information we refer to as "knowledge." Much of man's thinking about knowledge has been of theological and metaphysical nature, and it has answered few questions with any degree of assurance. Can it be said that all knowledge exists a priori in the universe - since God is omniscient, and since every physical law which we discover must have somehow "existed?" Or is knowledge man's artifact by which he seeks to introduce relative order into absolute non-order? Is it a creation of man's language? Can there be knowledge without language? We don't know, nor do we have the tools to answer these questions.

Information science seeks another approach, a quantitative one, and it poses questions which are not necessarily new but which information science has rendered more realistic and practical to answer. Among these are questions such as the size of man's knowledge and the rate of its growth.

An answer to the question "What is the amount of knowledge in the World?" is clearly of little practical importance to the individual; his information processing rates of a few dozen bits per second are too slow for him to store but a fraction of this knowledge during his entire earthly existence. If we are willing to admit, however, that there may be an occasion at some time to either appreciably extend man's lifetime (e.g., by travel at the speed of light) or to pass on knowledge "en bloc" to some extraterrestrial civilization having a high learning curve, such an inquiry need not be absolutely pointless.

It seems that beyond the simple instincts, man's ability to know and understand is dependent on language. Certainly his ability to communicate knowledge is language dependent. It is thus more precise to rephrase our question to read: "How much knowledge does mankind have?" and to expect that, barring the still remote possibility of identifying and decoding the content of human memory, we shall be largely limited in our measurement of knowledge to the measurement of recorded information.

Theoretically, we can readily measure the amount of information in the world by counting characters, marks and other symbols in records of all kinds; however, this amount of information is not equivalent to the totality of knowledge. Knowledge implies more than information in the sense that a message or a string of characters can have more than one meaning. Thus information theory (Shannon), in which the amount of information is a measure of the unexpectedness of its occurrence, is not yet useful in the measurement of knowledge, and the theory of semantic information (Bar-Hillel and Carnap) not yet sufficiently well developed. The measurement of information on a probabilistic basis has not yet developed a non-subjective measure of meaning and value, but in my opinion such a measure will be found eventually.

If the total amount of information in the world is not equivalent to the totality of knowledge, can we in the meantime determine whether it is larger or smaller?

If human knowledge is larger in amount than the total volume of information in the records of mankind, it is so because some or all the symbols and symbol combinations have more than a unique meaning, and because individuals bring to bear on this meaning their relevant and didfferent memories. In this instance, we again lack the necessary unit of measure to determine the difference.

If we for a moment disregard the plurality of meaning and the question of the role of grammar, the amount of knowledge can very well be less than the totality of information. Imagine one giant human processor which processes and stores every non-identical bit of information: the question of how its amount of knowledge differs from the amount of data becomes essentially a problem of information structures - the relations of content-carrying symbols. These relations, by means of which we should be able to separate, for instance, "old" knowledge from "new," are more subtle than the deceivingly simple task of identifying duplication; in fact, upon reflection you can see that we are harping on the problem of knowledge growth.

The process of knowledge growth is one of the most fascinating yet little understood information processes. We think it proceeds by certain mechanisms; those of addition, resolution, condensation and perhaps others. Knowledge is additive in the cumulative sense, it builds upon itself. The mechanism of resolution refers to a metamorphosis of information, it interacts with and is in turn modified by other information. The process of condensation appears to proceed via the mechanism of hierarchy (specific to general, concrete to abstract); the formulation of a law in science of the "new mathematics" are good examples of this mechanism at work.

Processes such as those of addition, condensation and resolution probably counteract each other in some measure - yet we do not know which one predominates, or under what conditions. Some very tentative results* indicate that the rate of knowledge increase is more nearly linear, contrary to the popular concept of an exponential growth; what doubles every ten or eight years is the number of recorded symbols, pages of paper, magazines, discs, records, pictures, TV shows, etc. but not the knowledge of man. On the other hand, one wonders whether the tentative evidence of the synergistic nature of systems, as it is being investigated by some cyberneticians, may also hold for the system of man's knowledge.

With a better understanding of the relations of concept-bearing symbols we shall be able to approach the question of the structure of knowledge. The payoff from this direction will be monumental. For instance, knowing that learning very much depends on the structure of the material leaned, an algorithm for optimum organization of subject contents will be feasible, with the result that man will learn much more efficiently.

I think it is not an exaggeration to expect that with the understanding of information structures we shall be approaching the possibility of automatic

---

* See, for example, A. D. Booth, "Mechanical Resolution of Linguistic Problems," in Kent and Taulbee, eds. Electronic Information Handling. Washington, D. C. , 1965, p. 49.

generation of new knowledge. As you are well aware, our use of information today is determined by the limitations of our systems for access to it. Today, the most prevalent system of access to information is via some descriptive tags, attached to it by man or machine; this mechanism is imprecise and its effectiveness therefore decreases with the increasing amount of information so tagged. A more complex mechanism, one which ignores the artificial packaging of information so tagged, is under development. However, both index and full-text searching lack precision. This lack is caused by the relativity of information, artificiality of its packaging (whether into books or sentences), and other factors; hence the imperfect results with mechanical translation, automatic abstracting and document retrieval.

Now if you accept a hypothesis of mine; that most "new" knowledge is generated by the application of existing or modified methods to new sets of problems, it is theoretically possible, once the structuring relations of information are understood - to generate new knowledge by information processing automata. My hypothesis identifies the method as the moving force of science and of knowledge growth, and it is not difficult to imagine information processing systems and mechanisms having an algorithmic capability of fostering an iteration of methods and problems. With a repertoire of logical and mathematical aids programmed into it, such systems would automatically seek out methods likely useful to the solution of given problems, and proceed to solve them using these aids.

CONCLUSION

I have singled out the problem of information structures as the cardinal problem of information science (and one of cardinal difficulty). Major progress in this direction will come only as a result of considerable work in many fields and directions. We must study the very nature and properties of symbols, signs and codes, and the manner in which they represent concepts. We must deal with the question and mechanism of human association, and the related problem of relevance. We must answer questions of the meaning of difference between and inconsistency of humans, and perhaps decide whether we wish to retain or overcome these characteristics. We must decide whether or not information theory must necessarily remain within a subjective framework of reference, and proceed to develop a measure of information value. We must study the concept of memory and its organization in man and automata. In these and all other endeavors, we should prudently and continually inquire into the differences between what is useful, possible, and desirable. The road is a long one, and the resources too precious and valuable to waste.

REFERENCE

1. Abelson, P. H.: Trends in Scientific Research. Science, Jan. 17, 1964, pp. 218-23.

DISTRIBUTION

INTERNAL

CC-P

DEP-T

DIR

E

I-RM-M

MA-PT (5)

MS-T (6)

MS-H

MS-IL (18)

MS-IP

PA (2)

R-DIR (2)

R-AS (5)

R-S

R-TO

R-AERO (5)

R-ASTR (5)

R-COMP (2)

R-ME

R-P & VE (5)

R-QUAL

R-RP-J
C. Swanson (10)

R-TEST (2)

I-DIR

EXTERNAL

National Aeronautics and Space
Administration
Ames Research Center
Moffett Field, California 94035 (2)

John F. Kennedy Space Center, NASA
Kennedy Space Center, Florida 32899 (2)

National Aeronautics and Space
Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771 (2)

National Aeronautics and Space
Administration
Langley Research Center
Langley Station
Hampton, Virginia 23365

National Aeronautics and Space
Administration
Lewis Research Center
21000 Brookpark Road
Cleveland, Ohio 44135 (2)

National Aeronautics and Space
Administration
Manned Spacecraft Center
Houston, Texas 77058 (2)

National Aeronautics and Space Administration
Wallops Station
Wallops Island, Virginia 23337 (2)

National Aeronautics and Space Administration
Western Operations Office
150 Pico Boulevard
Santa Monica, California 90496 (2)

National Aeronautics and Space Administration
Washington, D. C. 20546
Code:
ATSP
ATS
AFE
AFG
AFA
T (2)
M (5)
M-N (2)
S (5)
SP (2)
SC (2)
R (5)
RR (2)

Scientific and Technical Information Facility (25)
P. O. Box 33
College Park, Maryland 20740
Attn: NASA Rep. (S-AK/RKT)

Vladimir Slamecka
Georgia Institute of Technology
Atlanta, Georgia (25)